Supplementary information for

Lineage barcoding in mouse with homing CRISPR

Authors: Kathleen Leeper[1,2], Kian Kalhor[3], Andyna Vernet[4], Amanda Graveline[4], George M. Church[4,5], Prashant Mali[3], and Reza Kalhor[1,2,6,7]

Affiliations:

[1]*Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD, USA*

[2]*Center for Epigenetics, Johns Hopkins University School of Medicine, Baltimore, MD, USA*

[3]*Department of Bioengineering, University of California San Diego, La Jolla, CA, USA.*

[4]*Wyss Institute for Biologically Inspired Engineering at Harvard University, Boston, MA, USA*

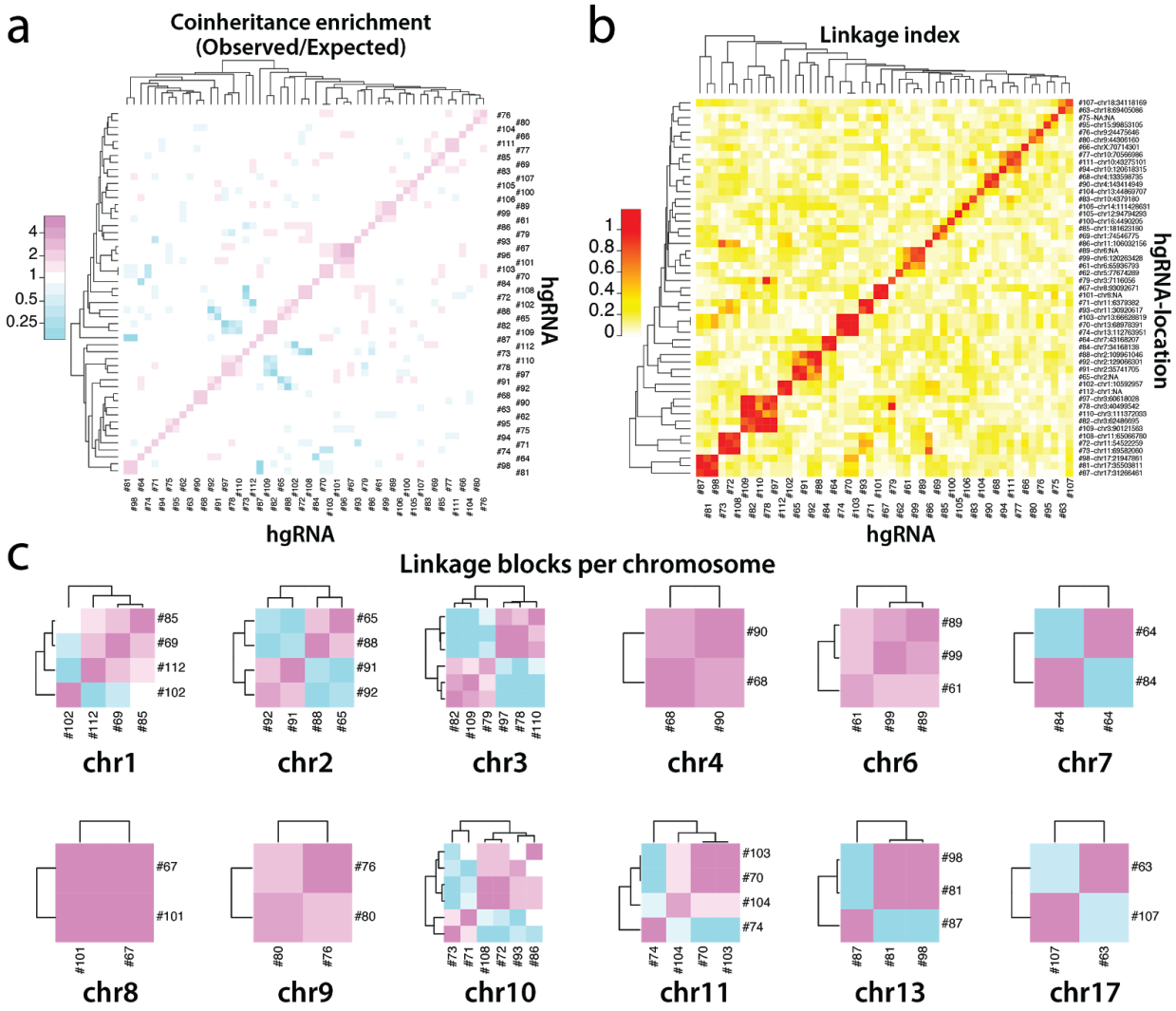
[5]*Department of Genetics, Harvard Medical School, Boston, MA, USA*

[6]*Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD, USA*

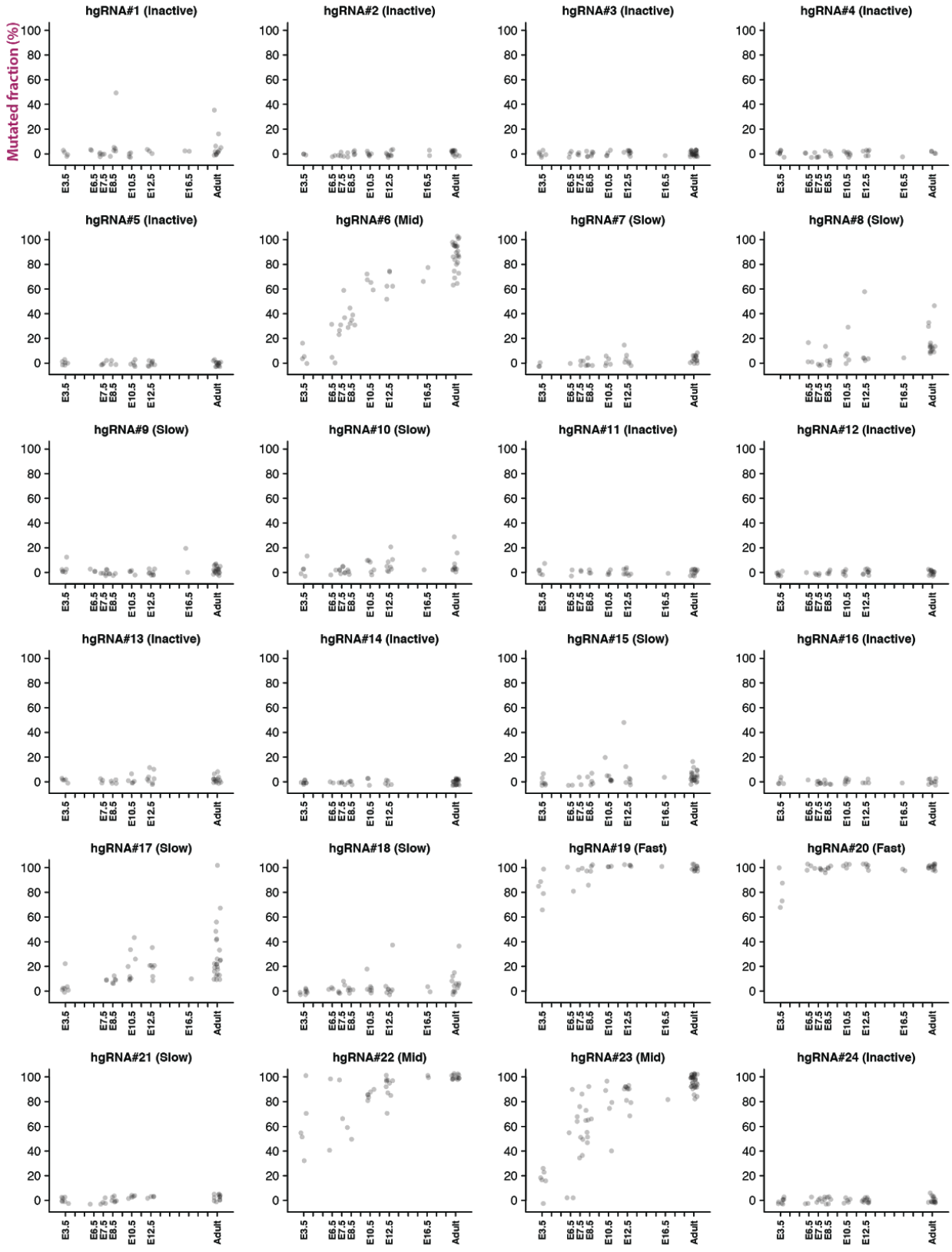
[7]*Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA*

Correspondence should be addressed to R.K. (kalhor@jhu.edu).

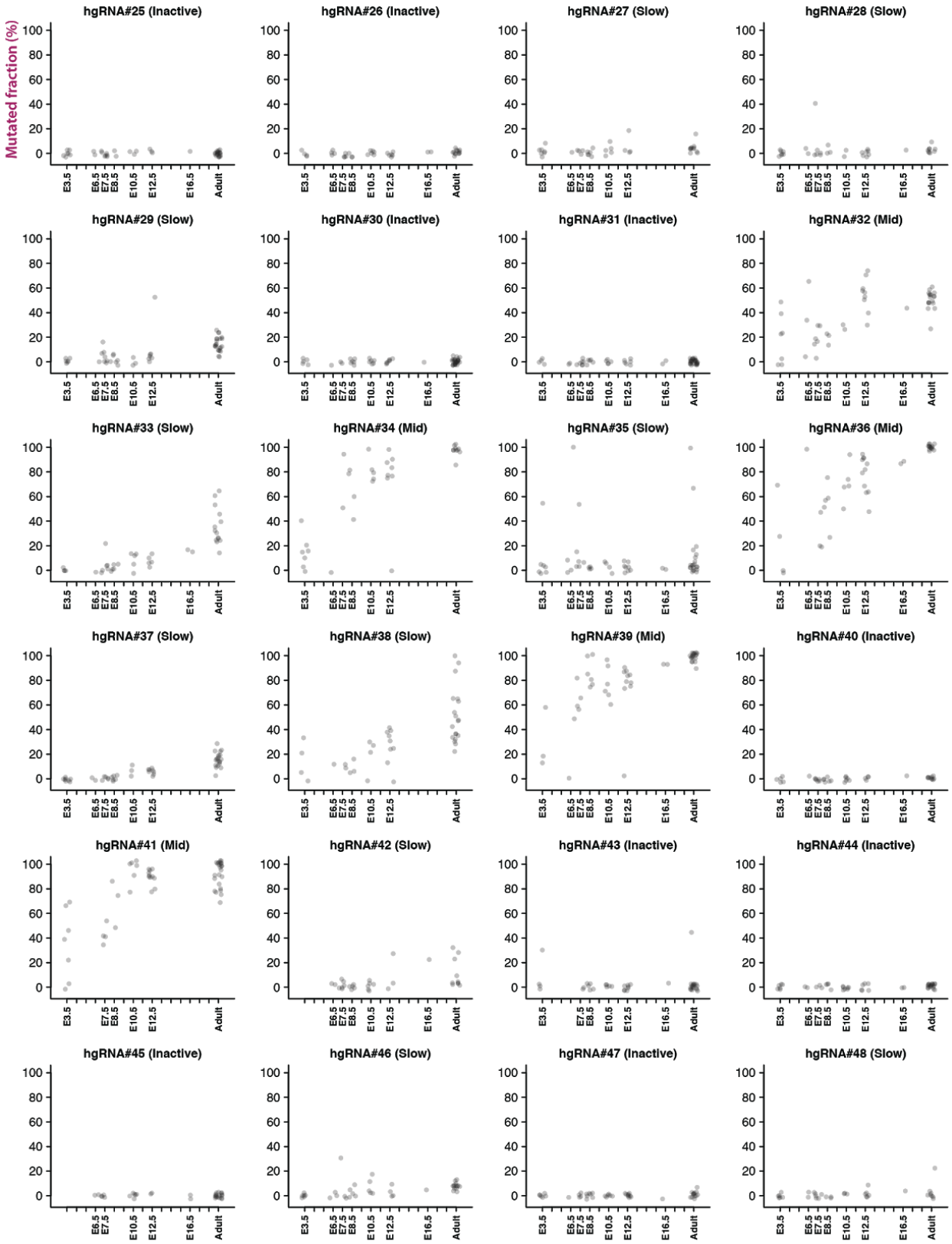
Supplementary Figures



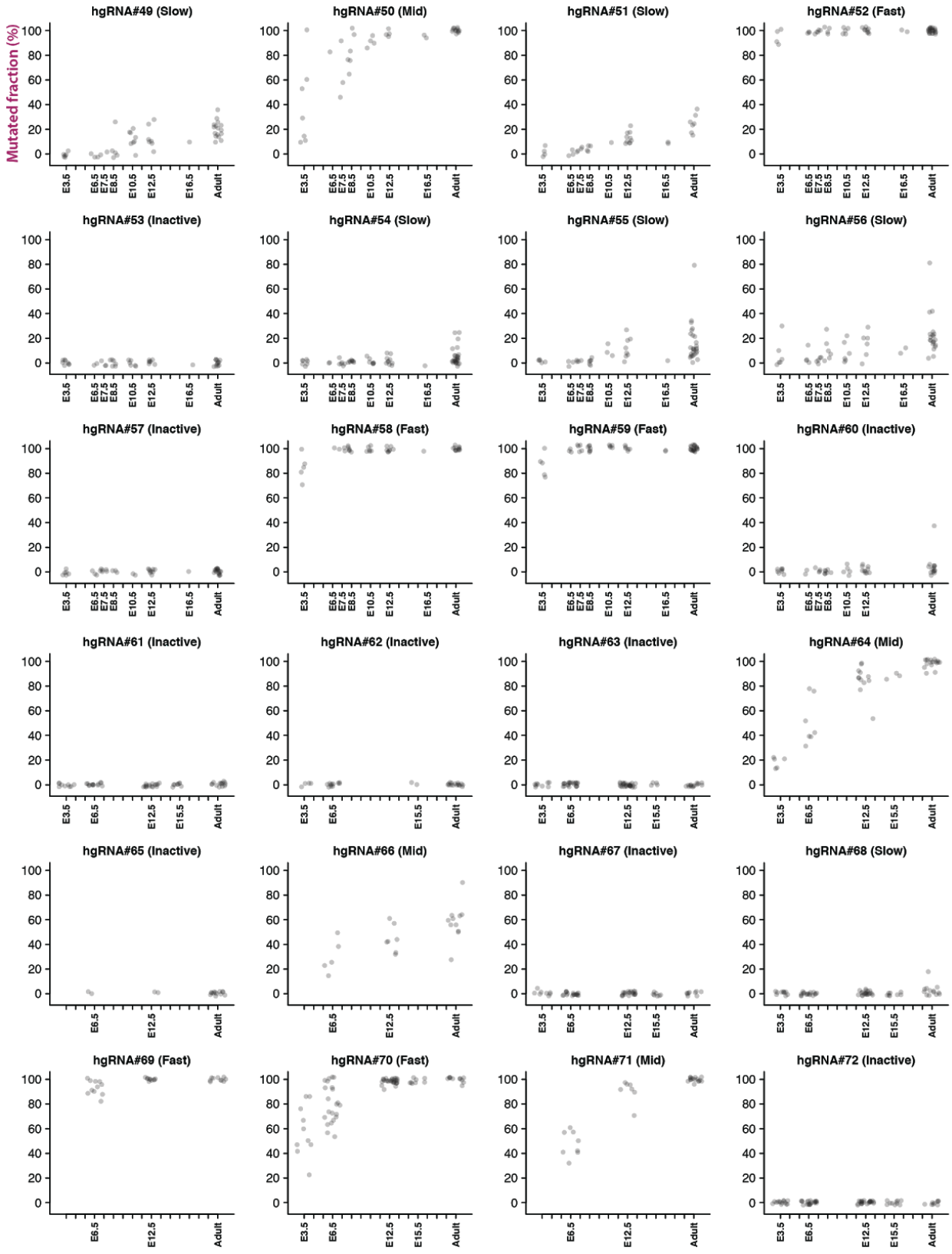
Supplementary Figure 1. Linkage disequilibrium of PB3 founder hgRNAs. **(a)** Coinheritance enrichment of all PB3 hgRNAs pairs, indicating elevated (purple color gradient) or diminished (blue color gradient) probabilities that two hgRNA are passed on to the same progeny. **(b)** Linkage index of all pairs for hgRNAs with uniquely identified genomic positions, indicating their possibility of residing less than 50 centimorgans on the same chromosome. **(c)** Coinheritance enrichment among linked hgRNAs on each chromosome. For the same data on PB7 founder hgRNAs, as well as derivation methods, see the original reference (Kalhor et al., Science, 2018).



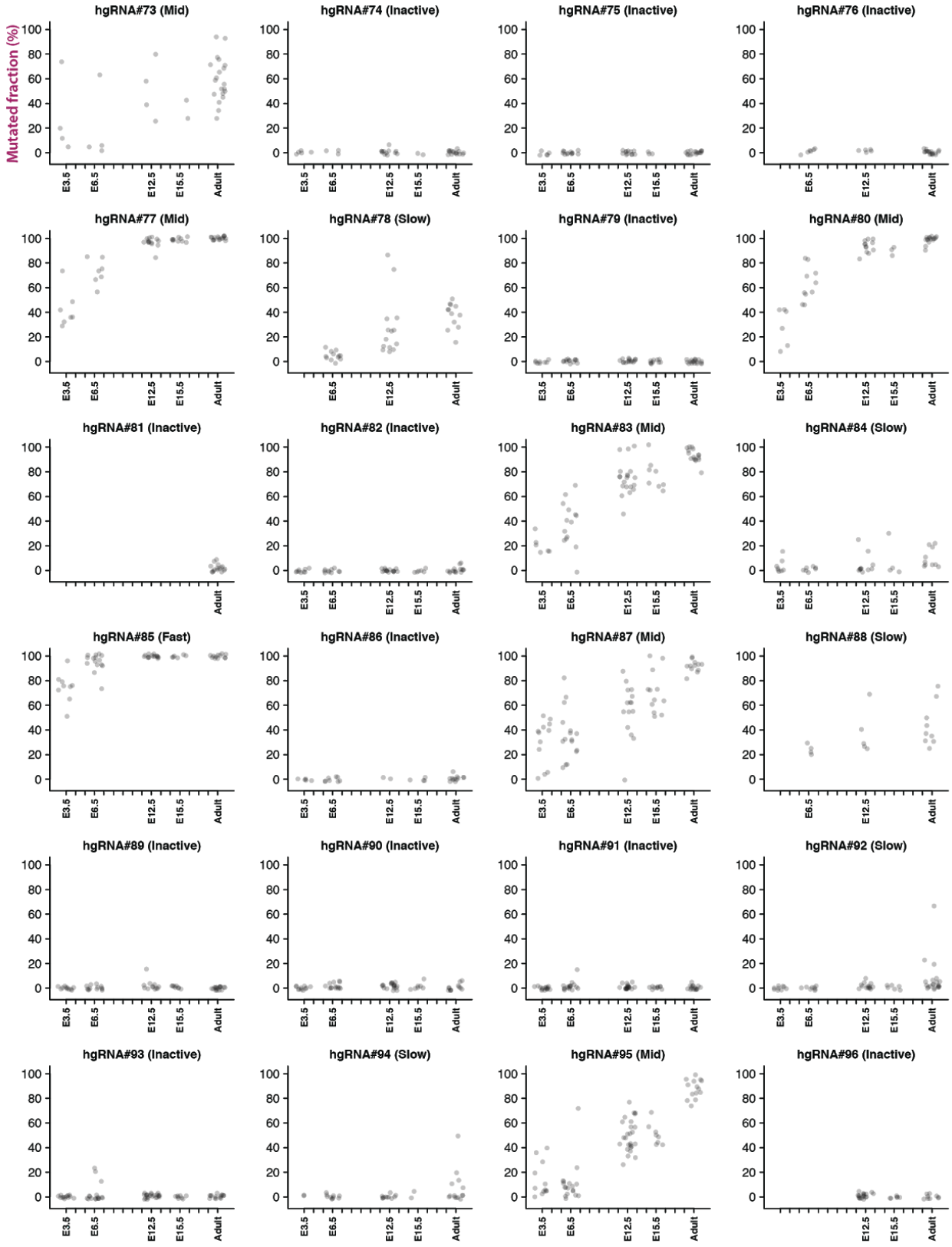
Supplementary Figure 2 (Cont'd).



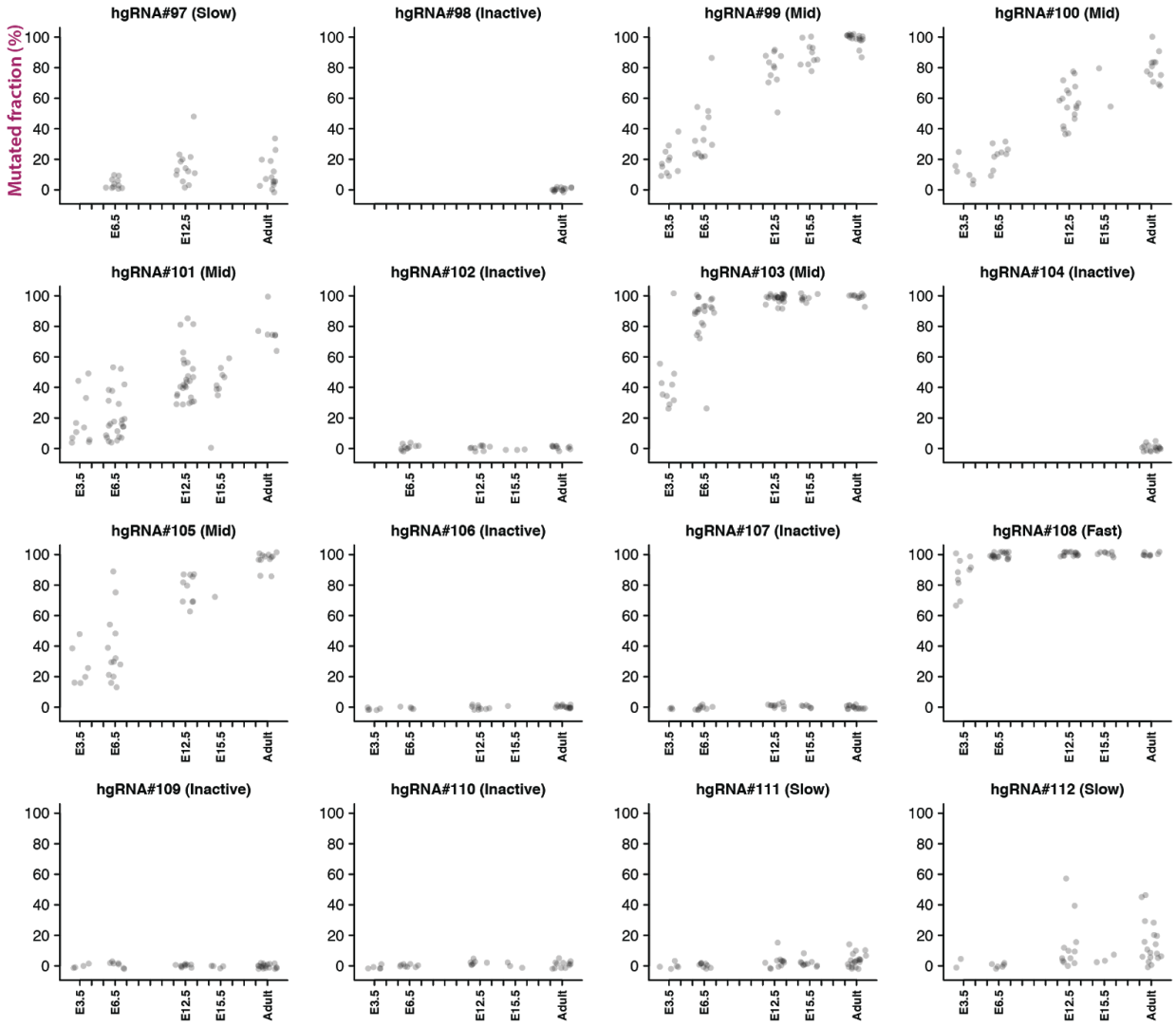
Supplementary Figure 2 (Cont'd).



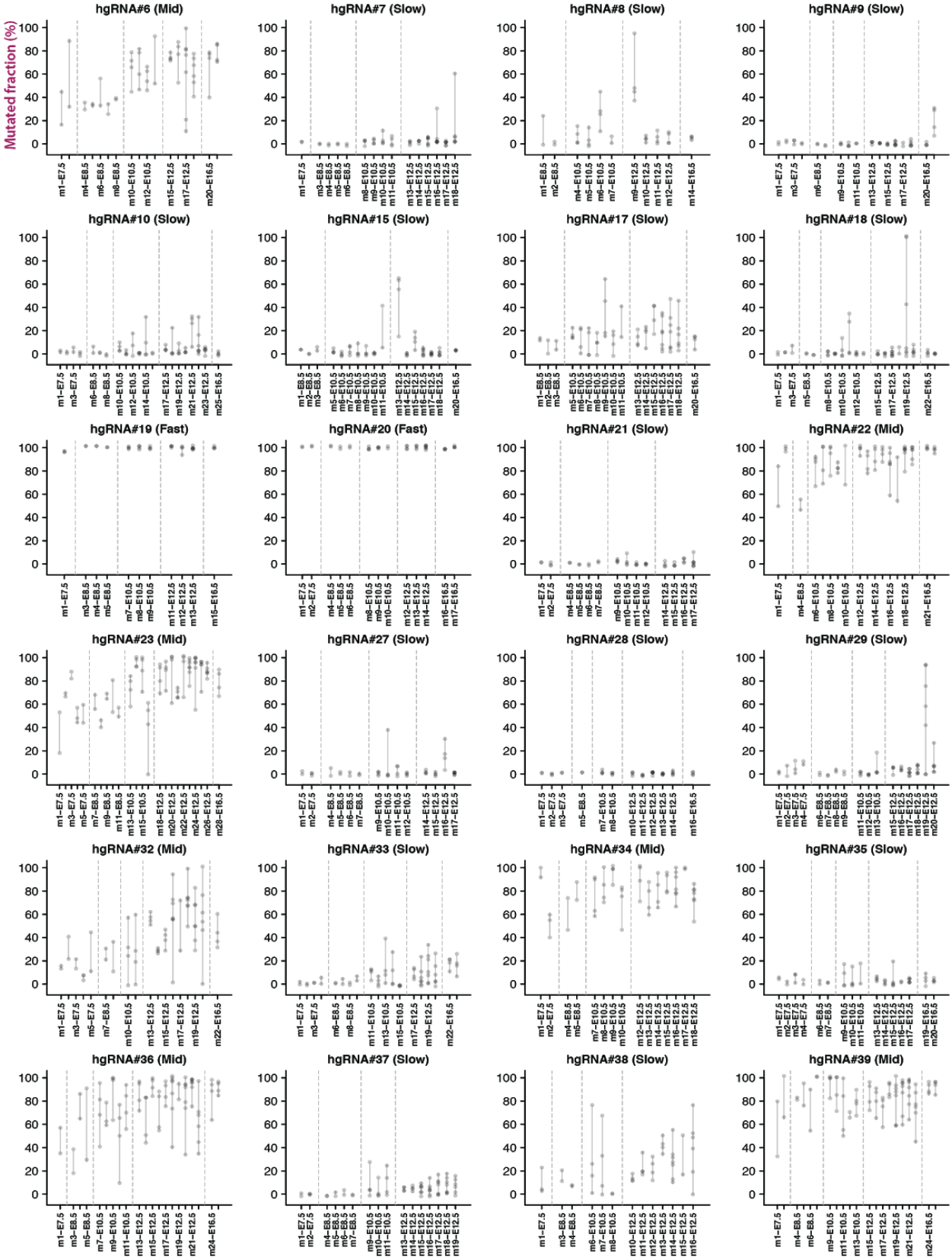
Supplementary Figure 2 (Cont'd).



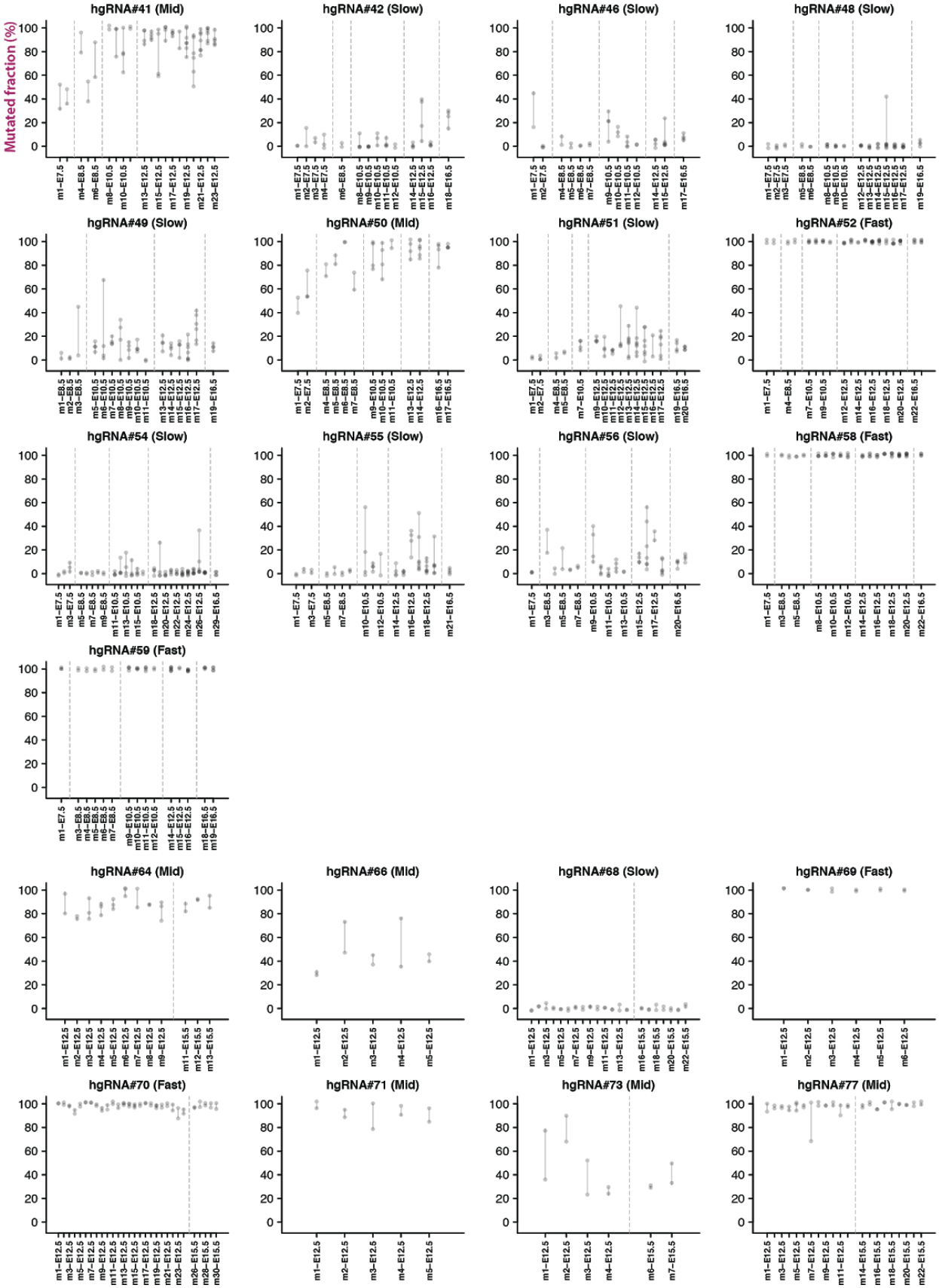
Supplementary Figure 2 (Cont'd).



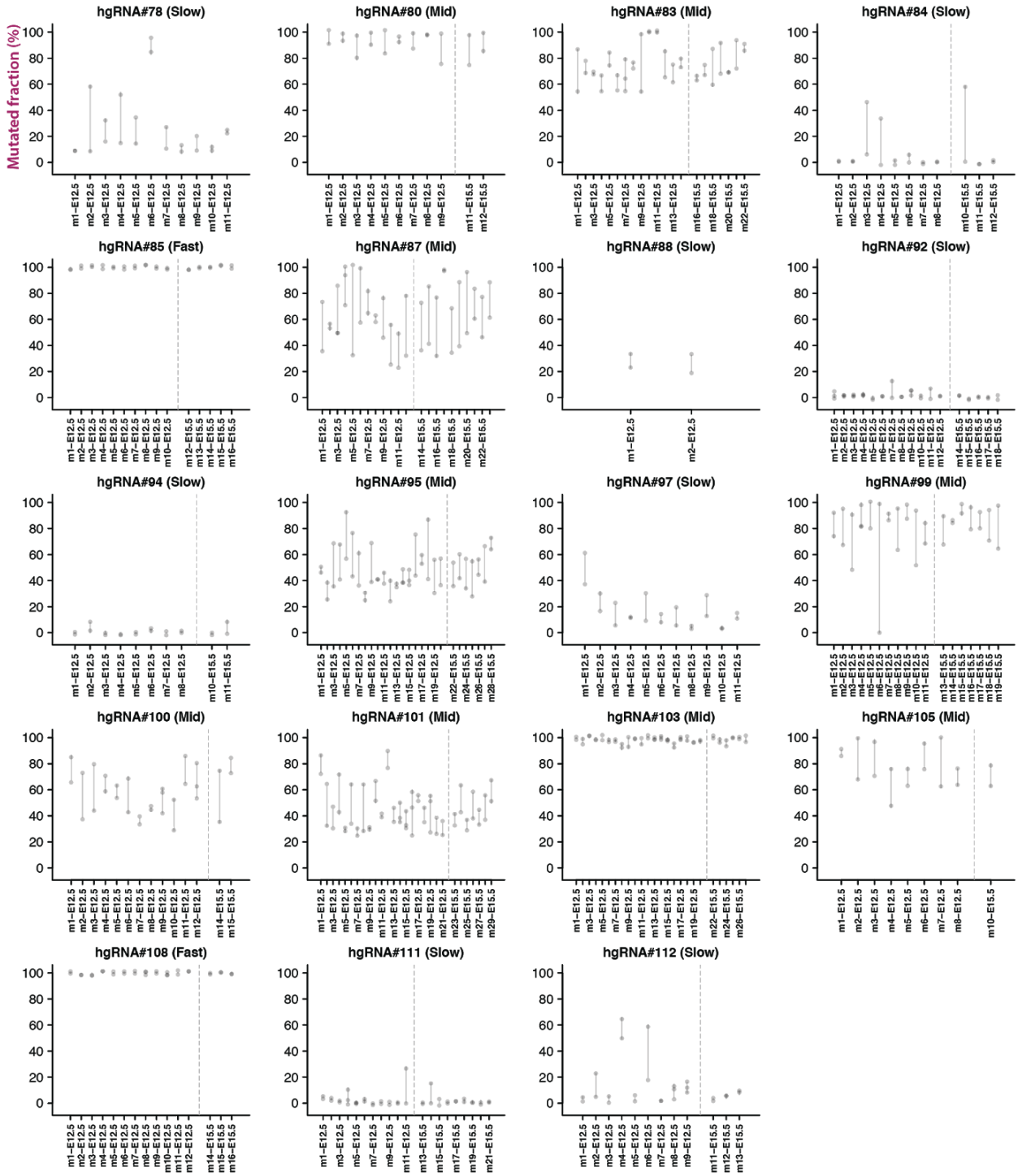
Supplementary Figure 2. Variation in hgRNA mutation levels in different barcoded mice. Each data point in each hgRNA's plot corresponds to the average mutated fraction of that hgRNA (Y-axis) in the samples taken from a single barcoded mouse at a given stage in embryogenesis (X-axis).



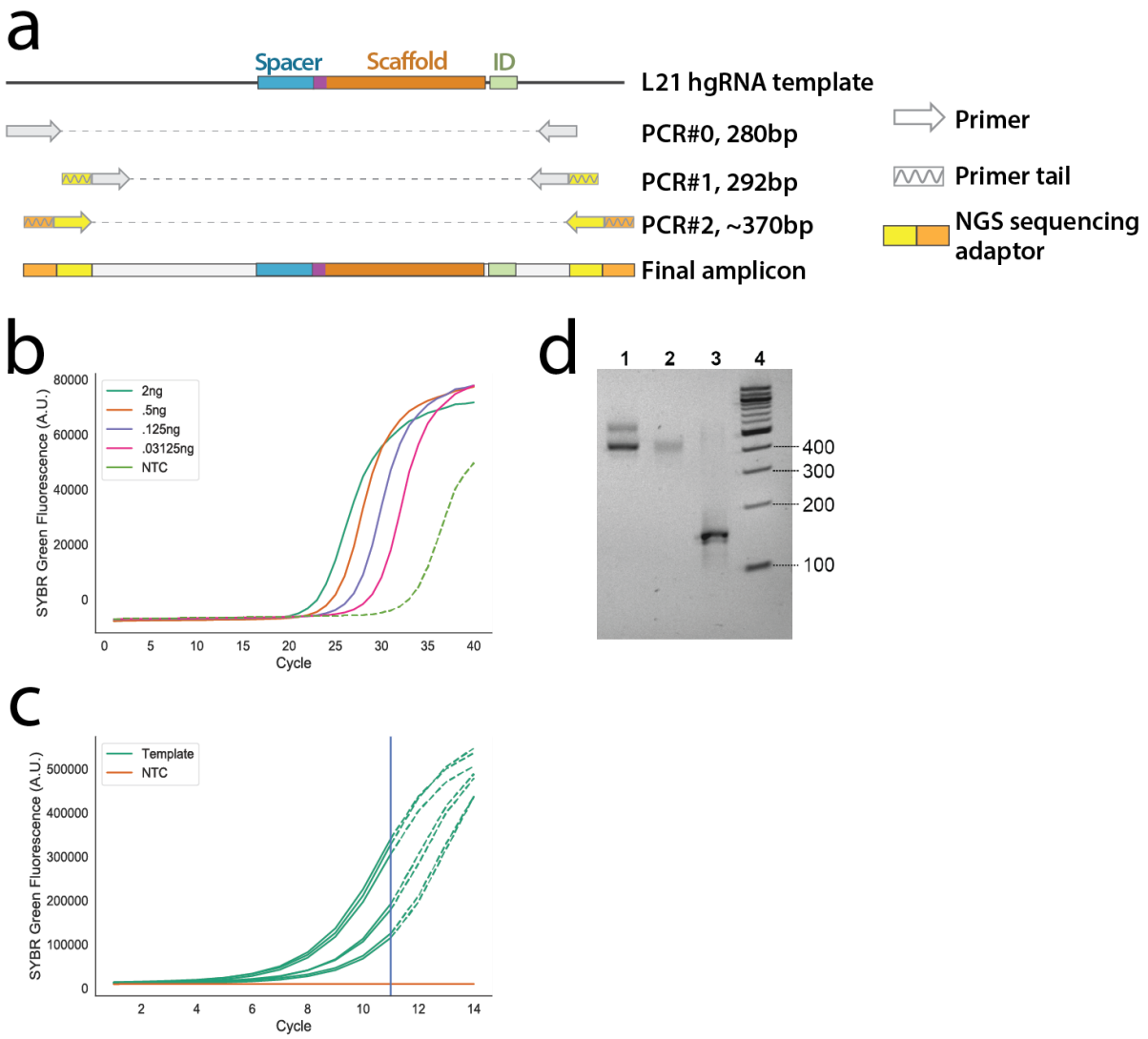
Supplementary Figure 3 (Cont'd).



Supplementary Figure 3 (Cont'd).



Supplementary Figure 3. Variation in hgRNA mutation levels within barcoded mice. Each column in each hgRNA's plot corresponds to a single barcoded mouse at a given stage of embryogenesis (X-axis); each dot within a column corresponds to the observed mutated fraction of that hgRNA (Y-axis) in a different sampled tissue. The line connecting the dots indicates range. Vertical dashed lines separate embryos in different stages.



Supplementary Figure 4. Amplification layout and expected outputs for real time PCR and library preparation. **(a)** Layout of hgRNA template in genomic DNA together with the binding site of primers for different PCR steps. **(b)** Representative standard curve for genomic DNA concentrations without pre-amplification. We observe consistent non-specific product formation in no-template controls (NTCs) above cycle 30 (dashed line), likely due to primer length, and suggest loading at least 0.2 ng of template such that all samples amplify well before these products form. **(c)** Representative SYBR Green data from a PCR#2 of 7 different samples and 1 NTC. Blue vertical line denotes an appropriate cutoff point; dashed lines represent overamplification phase. **(d)** 4% agarose gel of pooled and purified libraries. Lane 1: over-amplified reactions with both correctly sized band and heteroduplex formation, Lane 2: correctly amplified reactions, Lane 3: non-specific PCR#1 products carried through to PCR#2. Lane 4: 100 bp ladder, bands at 100bp intervals.

Supplementary Tables

Supplementary Table 1 (provided as a separate file). List of all hgRNAs in the MARC1 founder males. For each hgRNA, its number, ID sequence, spacer sequence, TSS to PAM length, spacer length, observed inheritance probability, chromosomal location, and confidence level of the chromosomal location are given. A “+” sign in the location column indicates that the hgRNA is transcribed in the same direction as the positive strand of the mm10 reference sequence.

Supplementary Table 2. Sample numbers for various developmental stages. Each sample per embryo represents the total number of unique tissue samples within that embryo. Values indicate the number of embryos sampled at each stage. When multiple samples were obtained from an embryo, different regions of both embryonic and extraembryonic tissues were targeted. For adult mice, samples correspond to P18-P25 stage and comprise ear notches. For E3.5 and E6.5, all samples are whole conceptus. For E7.5 samples are either whole conceptus or conceptus divided in 2 or 3 segments along its anterior–posterior axis. For E8.5 through E16.5, the placenta, parietal yolk sac, visceral yolk sac, head, heart, limb, or tail, or a combination thereof were collected. For these E8.5–E16.5 samples, cases with more than a single sample represent at least one sample from embryonic and another from extraembryonic tissues.

Line	Samples per embryo	Developmental Stage										Total Mice	Total Samples
		E3.5	E6.5	E7.5	E8.5	E10.5	E12.5	E14.5	E15.5	E16.5	Adult		
PB7	1	11	6	3	3	1	0	0	0	0	42	66	66
	2	0	0	4	6	1	0	0	0	0	0	11	22
	3	0	0	1	0	0	0	0	0	0	0	1	3
	4	0	0	0	0	7	6	0	0	2	0	15	60
	5	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	2	0	0	0	0	2	12
	7	0	0	0	0	0	4	0	0	0	0	4	28
	total	**11**	**6**	**8**	**9**	**9**	**12**	**0**	**0**	**2**	**42**	**99**	**191**
PB3	1	18	31	0	0	0	6	0	3	0	29	87	87
	2	0	0	0	0	0	25	0	12	0	0	37	74
	3	0	0	0	0	0	6	0	0	0	0	6	18
	4	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	0	0	0	0	0	0
	total	**18**	**31**	**0**	**0**	**0**	**37**	**0**	**15**	**0**	**29**	**130**	**179**

Supplementary Table 3 (provided as a separate file). Average observed mutation level of each hgRNA during different stages of development. Values indicate percent mutated while “NA” indicates values that were not measured. Functional category based on this mutation level is also shown. Barcodes with less than 2% maximum mutation rate in all samples were assigned as “inactive”, those with 2-50% maximum mutation rate in all samples were assigned as “slow”, those with more than 90% average mutation frequency at any point before E9 were assigned as “fast”, and the rest were assigned as “mid.”

Supplementary Table 4 (provided as a separate file). For each mutant spacer allele of each hgRNA, its sequence, occurrence probability (fraction of mice with that hgRNA which contained the allele), occurrence (number of mice out of total with that hgRNA which contained the allele), and median abundance (median of its abundance fraction in all mice in which it was observed) are shown.

Supplementary Table 5 (provided as a separate file). Example genotyping output of the analysis pipeline, corresponding to a mutated sample. A [*sampleName*]_genotypes.txt file is produced for each sample It provides each identifier (# and ID) and the most commonly-observed associated spacer sequence (SP), forming an identifier–spacer pair; the absolute count of observations of this pair (count); the percent of this pair relative to all observed pairs in a sample; and the percent of this pair relative to all pairs with that identifier.

Supplementary Table 6 (provided as a separate file). Example filtered pairs output of the analysis pipeline on a mutated sample. A [*sample*]_filteredpairs.txt file, produced for each sample, contains information on mutation levels and all observed pairs. This table contains the observed identifier with a corresponding observed spacer, the total observation count for that pair, the percent of that pair within all observations of that identifier, and the percent of that pair within all sample observations.

Supplementary Table 7 (provided as a separate file). Suggested primer sets for specific amplification of each hgRNA in non-barcoded mice. Each pair is expected to produce an amplicon of 141–155 bp in length. Using each primer pair in a standard PCR on genomic DNA of MARC1 mice can establish the presence of the targeted hgRNA if an amplicon of the correct size appears. Absence of the correct amplicon in the PCR product would indicate absence of the hgRNA. These primer sets are not experimentally validated and may fail to amplify mutated hgRNA sequences as they depend on the spacer region to bind.

Supplementary Table 8 (provided as a separate file). Table of observed mutant allele frequencies in each sample in Figure 5. Rows correspond to samples and columns to each of the mutant spacer alleles that were observed. Alleles with less than 0.25% abundance in each sample are not included. parSP indicates spacer sequence observed in founder, mutSP indicates mutant spacer sequences observed in samples under analysis.